Opportunities of collaboration between biostatisticians and pre-clinical scientists: a literature review

Márcio Augusto Diniz, PhD

Co-Director, Biostatistics and Clinical Informatics Core at Tisch Cancer Institute

Associate Professor, Department of Population Health Science and Policy

Institute of Healthcare Delivery Science

Icahn School of Medicine at Mount Sinai

Reproducibility/Replicability crisis in science

Reproducibility/Replicability in Science

Published: 28 March 2012

Drug development

Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Nature **483**, 531–533 (2012) | Cite this article

- Clinical trials in oncology have the highest failure rate compared with other therapeutic areas;
- The quality of published preclinical data plays a central role as drug development relies heavily on the literature, especially with regards to new targets and biology;
- Scientists at Amgen tried to replicate 53 landmark preclinical studies;
- Only 11% (6) out of 53 studies were replicated.

REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

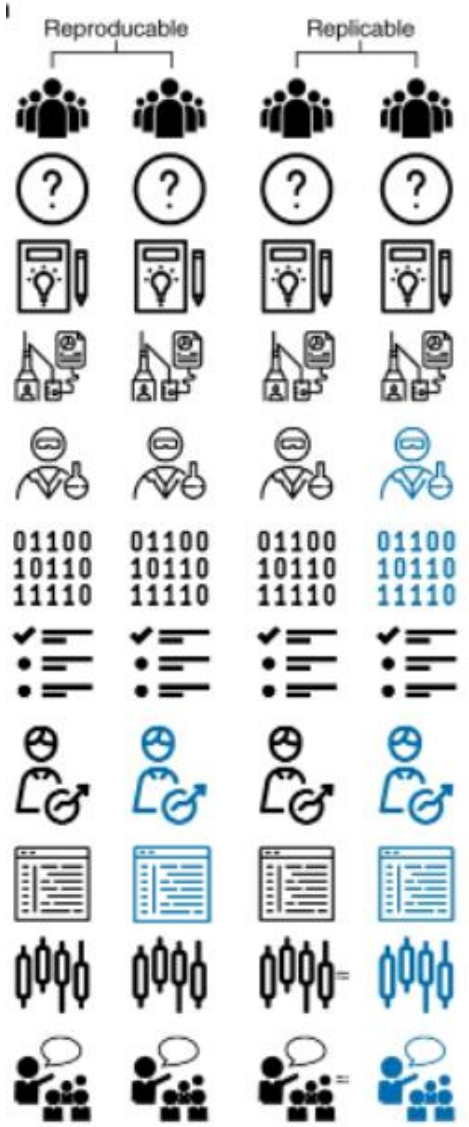
Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.
*Source of citations: Google Scholar, May 2011.

Begley CG, Ellis LM. Raise standards for preclinical cancer research. Nature. 2012 Mar;483(7391):531-3

Reproducibility vs Replicability

- When comparing two studies, possible sources of differences are:



Patil P, Peng RD, Leek JT. A visual tool for defining reproducibility and replicability. Nature human behaviour. 2019 Jul;3(7):650-2.

Reporting

OPEN ACCESS Freely available online

PLoS one

Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals

Carol Kilkenny[1]*, Nick Parsons[2], Ed Kadyszewski[3], Michael F. W. Festing[4], Innes C. Cuthill[5], Derek Fry[6], Jane Hutton[7], Douglas G. Altman[8]

- Kilkenny and colleagues reviewed 271 preclinical studies;
    - Only 59% stated the hypothesis or objective of the study and the number and characteristics of the animal used in the experiments;
    - 87% did not use randomization, 86% did not use blinding in their experiments;
    - Only 70% used statistical methods described in their methods and presented results with a measure of error such as standard deviation.

Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PloS one. 2009 Nov 30;4(11):e7824.

Reporting

- ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines were proposed in 2010;

- ARRIVE has 17 set of items that can be divided under sections of a paper:
  - Introduction: Title, Abstract, Objectives, Ethical Statement;
  - Method: Study Design, Experimental Procedures, Experimental Animals, House and Husbandry, Sample Size, allocating Animals, Experimental Outcomes, Statistical Methods;
  - Results: Baseline Data, Outcomes and Estimation, Adverse Events
  - Discussion: Funding

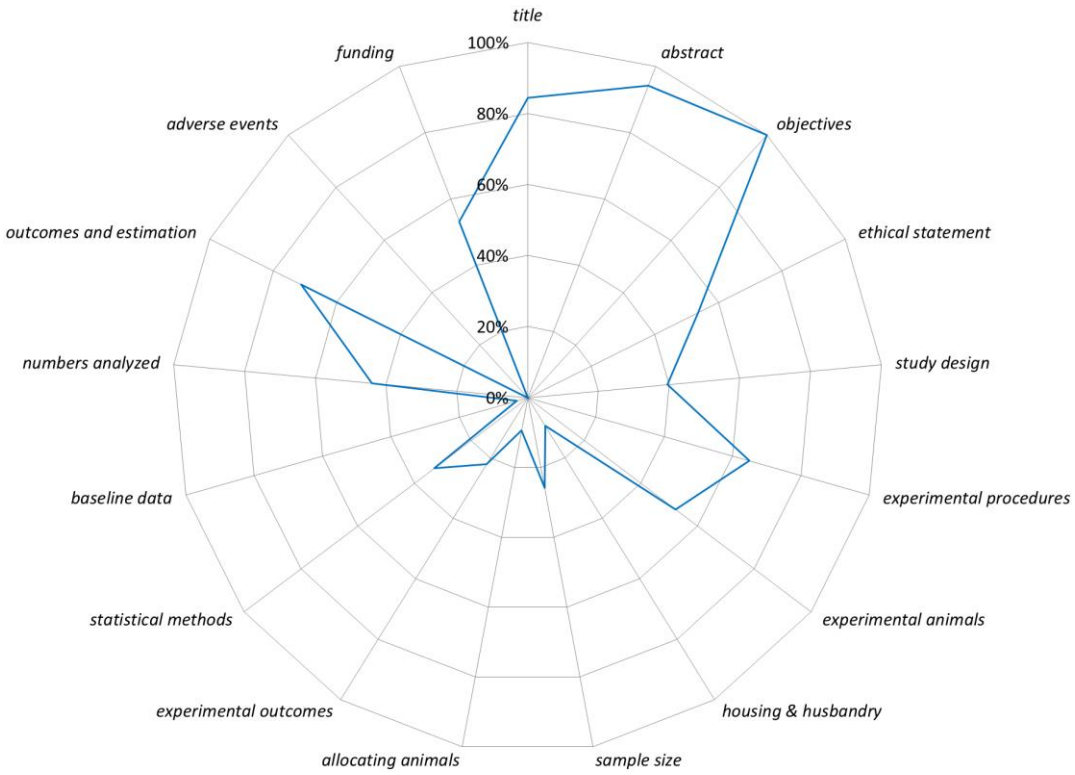
Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. Journal of Pharmacology and Pharmacotherapeutics. 2010 Dec;1(2):94-9.

Reporting

RESEARCH ARTICLE

The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research

Marc T. Avey[1,2]*, David Moher[1,3], Katrina J. Sullivan[1], Dean Fergusson[1], Gilly Griffin[1], Jeremy M. Grimshaw[1,4], Brian Hutton[1,3], Manoj M. Lalu[1,7], Malcolm Macleod[5], John Marshall[6], Shirley H. J. Mei[7], Michael Rudnicki[7], Duncan J. Stewart[7,8], Alexis F. Turgeon[9,10], Lauralyn McIntyre[1,11], Canadian Critical Care Translational Biology Group[¶]



- After 6 years that ARRIVE guidelines were proposed, Avey et al. reviewed 47 preclinical studies;
  - Adequate reporting of items from the Methods Section ranged from 9% (allocating animals to experimental groups, housing and husbandry) to 65% (experimental procedures);
  - Adequate reporting of items from the Results Section ranged from 0% (adverse events) to 71% (outcomes and estimation).

Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, Hutton B, Lalu MM, Macleod M, Marshall J, Mei SH. The devil is in the details: incomplete reporting in preclinical animal research. PloS one. 2016 Nov 17;11(11):e0166733

Reproducibility/Replicability in Science

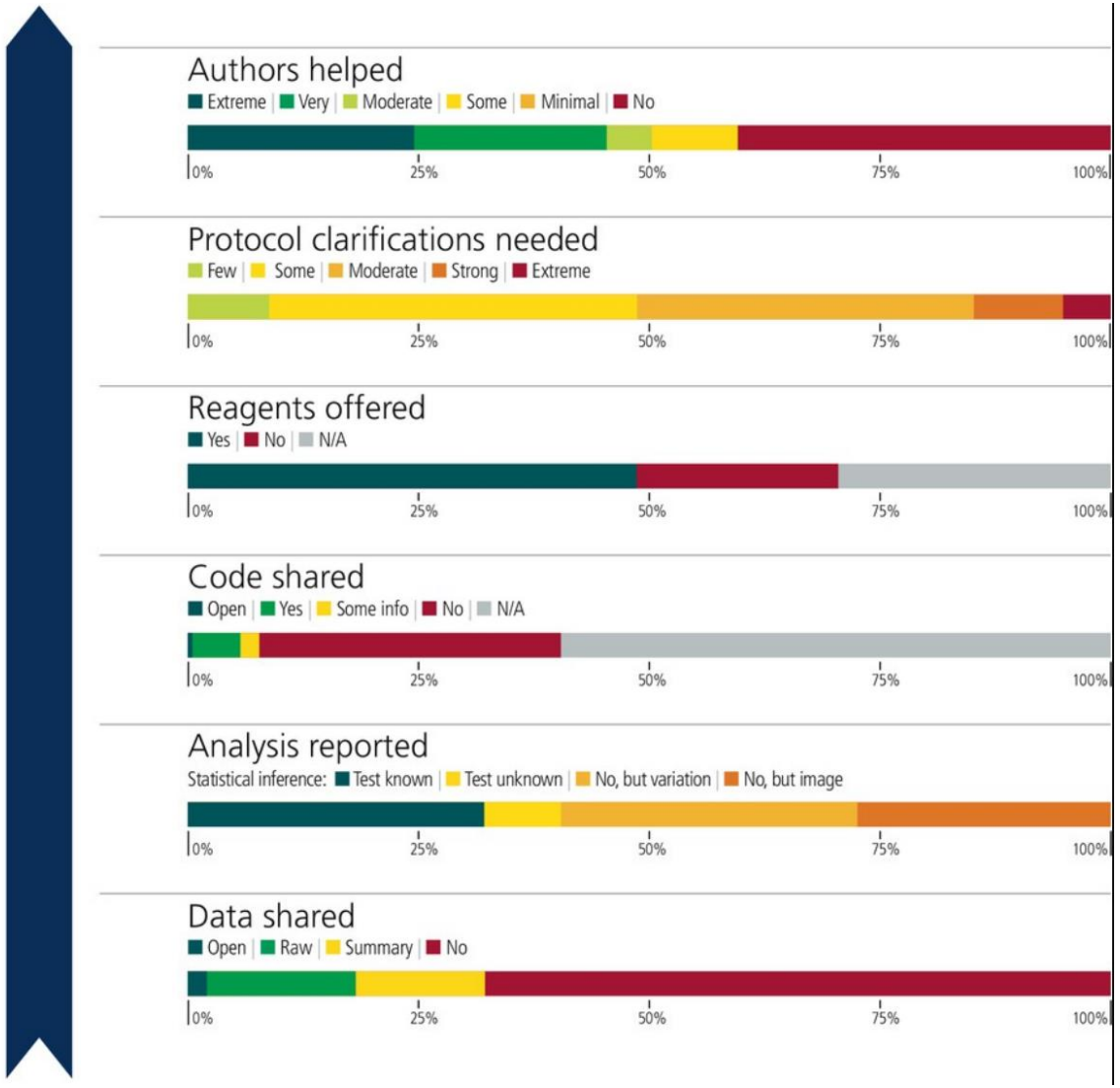
eLife FEATURE ARTICLE

REPRODUCIBILITY IN CANCER BIOLOGY

Challenges for assessing replicability in preclinical cancer biology

TIMOTHY M ERRINGTON*, ALEXANDRIA DENIS†, NICOLE PERFITO‡, ELIZABETH IORNS AND BRIAN A NOSEK

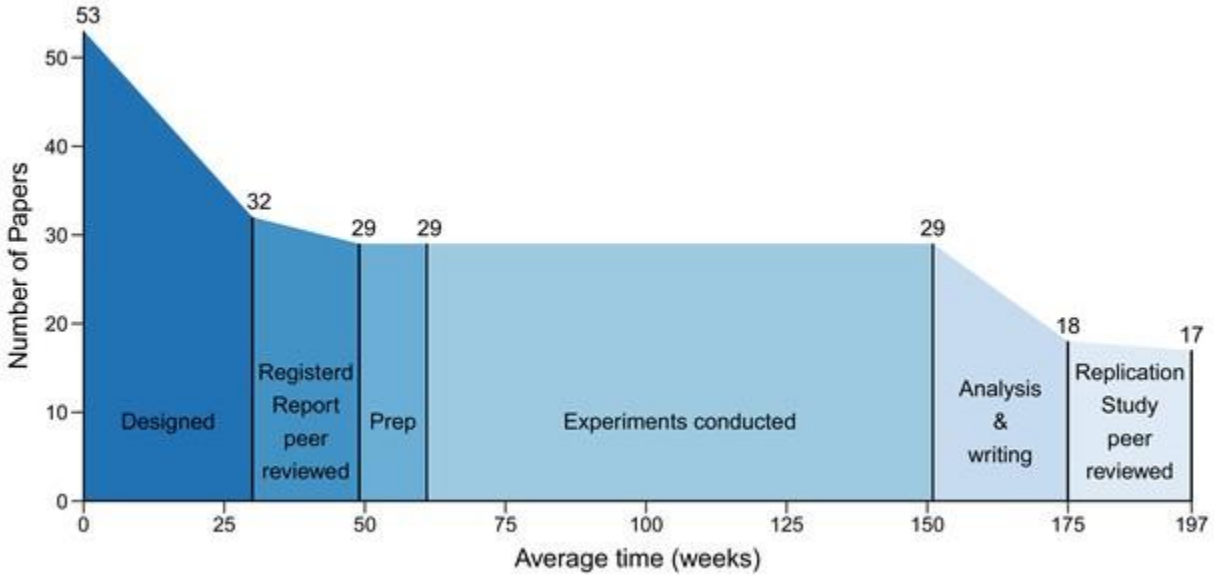
- Errington et al. tried to replicate 193 experiments from 53 high-impact papers as part of the project Reproducibility Project: Cancer Biology | Collections | eLife (elifesciences.org);

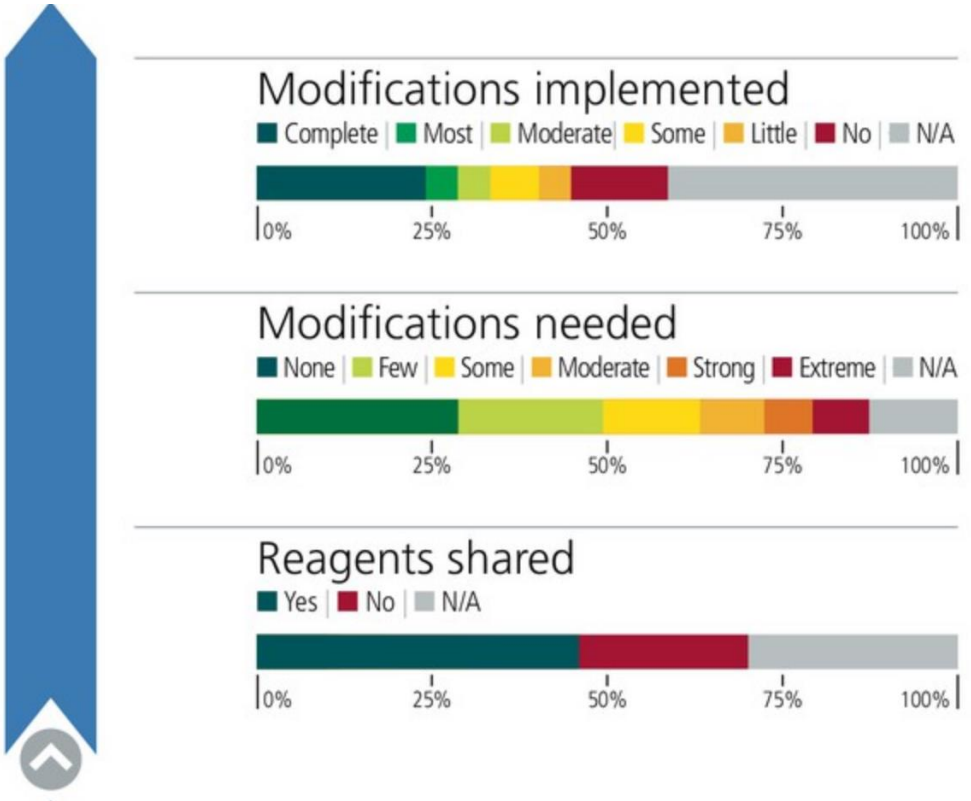


Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Reproducibility in cancer biology: challenges for assessing replicability in preclinical cancer biology. Elife. 2021 Dec 7;10:e67995.

Reproducibility/Replicability in Science

- Errington et al. initiated 87 experiments from 29 papers, but only completed 50 of them from 18 papers.





Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Reproducibility in cancer biology: challenges for assessing replicability in preclinical cancer biology. Elife. 2021 Dec 7;10:e67995.

Reporting

PLOS BIOLOGY

COMMUNITY PAGE

Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0

Nathalie Percie du Sert[1]*, Amrita Ahluwalia[2,3], Sabina Alam[4], Marc T. Avey[5], Monya Baker[6], William J. Browne[7], Alejandra Clark[8], Innes C. Cuthill[9], Ulrich Dirnagl[10], Michael Emerson[11], Paul Garner[12], Stephen T. Holgate[13], David W. Howells[14], Viki Hurst[1], Natasha A. Karp[15], Stanley E. Lazic[16], Katie Lidster[1], Catriona J. MacCallum[17], Malcolm Macleod[18], Esther J. Pearl[1], Ole H. Petersen[19], Frances Rawle[20], Penny Reynolds[21], Kieron Rooney[22], Emily S. Sena[18], Shai D. Silberberg[23], Thomas Steckler[24], Hanno Würbel[25]

Box 2. ARRIVE Essential 10

1. Study design
2. Sample size
3. Inclusion and exclusion criteria
4. Randomisation
5. Blinding
6. Outcome measures
7. Statistical methods
8. Experimental animals
9. Experimental procedures
10. Results

Box 6. ARRIVE Recommended Set

1. Abstract
2. Background
3. Objectives
4. Ethical statement
5. Housing and husbandry
6. Animal care and monitoring
7. Interpretation/scientific implications
8. Generalisability/translation
9. Protocol registration
10. Data access
11. Declaration of interests

Du Sert NP, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. PLoS biology. 2020 Jul 14;18(7):e3000411.

What have I learned from interacting with preclinical scientists working on research in stroke?

Reproducibility/Replicability in Stroke research

A call for transparent reporting to optimize the predictive value of preclinical research

Story C. Landis, Susan G. Amara, Khusru Asadullah, Chris P. Austin, Robi Blumenstein, Eileen W. Bradley, Ronald G. Crystal, Robert B. Darnell, Robert J. Ferrante, Howard Fillit, Robert Finkelstein, Marc Fisher, Howard E. Gendelman, Robert M. Golub, John L. Goudreau, Robert A. Gross, Amelie K. Gubitz, Sharon E. Hesterlee, David W. Howells, John Huguenard, Katrina Kelner, Walter Koroshetz, Dimitri Krainc, Stanley E. Lazic, ... Shai D. Silberberg + Show authors

Nature **490**, 187–191 (2012) | Cite this article

- The US National Institute of Neurological Disorders and Stroke convened major stakeholders in June 2012 to discuss how to improve the methodological reporting of animal studies in grant applications and publications.

Schmidt-Pogoda A, Bonberg N, Koecke MH, Strecker JK, Wellmann J, Bruckmann NM, Beuker C, Schäbitz WR, Meuth SG, Wiendl H, Minnerup H. Why most acute stroke studies are positive in animals but not in patients: a systematic comparison of preclinical, early phase, and phase 3 clinical trials of neuroprotective agents. Annals of neurology. 2020 Jan;87(1):40-51.

Reproducibility/Replicability in Stroke research

Why Most Acute Stroke Studies Are Positive in Animals but Not in Patients: A Systematic Comparison of Preclinical, Early Phase, and Phase 3 Clinical Trials of Neuroprotective Agents

Antje Schmidt-Pogoda, MD,[1,†] Nadine Bonberg, MSc,[2,†] Mailin Hannah Marie Koecke,[1] Jan-Kolja Strecker, PhD,[1] Jürgen Wellmann, PhD,[2] Nils-Martin Bruckmann, MD,[1] Carolin Beuker, MD,[1] Wolf-Rüdiger Schäbitz, MD,[3] Sven G. Meuth, MD, PhD,[1] Heinz Wiendl, MD,[1] Heike Minnerup, MD, MSc,[2] and Jens Minnerup, MD[1]

- Pivotal study design differences between experimental studies and clinical trials, including different primary end points and time to treatment, publication bias, neglected quality criteria and low power, contribute to the stepwise efficacy decline of stroke treatments from experimental studies to phase 3 clinical trials.

Schmidt-Pogoda A, Bonberg N, Koecke MH, Strecker JK, Wellmann J, Bruckmann NM, Beuker C, Schäbitz WR, Meuth SG, Wiendl H, Minnerup H. Why most acute stroke studies are positive in animals but not in patients: a systematic comparison of preclinical, early phase, and phase 3 clinical trials of neuroprotective agents. Annals of neurology. 2020 Jan;87(1):40-51.

The Stroke Preclinical Assessment Network

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

STROKE

A multi-laboratory preclinical trial in rodents to assess treatment candidates for acute ischemic stroke

Patrick D. Lyden[1,2*], Márcio A. Diniz[3], Francesca Bosetti[4], Jessica Lamb[1], Karisma A. Nagarkatti[1], André Rogatko[3], Sungjin Kim[3], Ryan P. Cabeen[5], James I. Koenig[4], Kazi Akhter[6], Ali S. Arbab[7], Brooklyn D. Avery[8], Hannah E. Beatty[9], Adnan Bibic[6], Suyi Cao[8], Ligia Simoes Braga Boisserand[9], Angel Chamorro[10,11], Anjali Chauhan[12], Sebastian Diaz-Perez[13], Krishnan Dhandapani[14], Nirav Dhanesha[15], Andrew Goh[12], Alison L. Herman[9], Fahmeed Hyder[16,17], Takahiko Imai[18], Conor W. Johnson[9], Mohammad B. Khan[19], Pradip Kamat[19], Senthilkumar S. Karuppagounder[20], Mariia Kumskova[15], Jelena M. Mihailovic[16], Joseph B. Mandeville[18], Andreia Morais[18], Rakesh B. Patel[15], Basavaraju G. Sanganahalli[16], Cameron Smith[19], Yanrong Shi[8], Brijesh Sutariya[15], Daniel Thedens[21], Tao Qin[18], Sofia E. Velazquez[9,13], Jaroslaw Aronowski[12], Cenk Ayata[22], Anil K. Chauhan[15], Enrique C. Leira[10,23,24], David C. Hess[19], Raymond C. Koehler[8], Louise D. McCullough[12], Lauren H. Sansing[9,13]

- Six interventions were selected to be tested in a multi-laboratory pre-clinical trial;
- Six independent research laboratories performed a standard focal cerebral ischemic insult in animals divided in five animal models: young mice, young rats, aging mice, mice with diet-induced obesity, and spontaneously hypertensive rats;
- Equal numbers of males and females;
- 2645 animals were enrolled throughout four stages with one drug selected at the end of the trial.

Lyden PD, Diniz MA, Bosetti F, Lamb J, Nagarkatti KA, Rogatko A, Kim S, Cabeen RP, Koenig JI, Akhter K, Arbab AS. A multi-laboratory preclinical trial in rodents to assess treatment candidates for acute ischemic stroke. Science translational medicine. 2023 Sep 20;15(714):eadg8656.

The Stroke Preclinical Assessment Network

Good practices:

- Blinding;
- Randomization;
- Allocation concealment
- Stratification by sex;
- Introduction of controlled variability;
- Adaptive sample sizes;
- Reporting followed ARRIVE 2.0 guidelines.

Opportunities of improvement:

- Several drugs might have failed due lack of adequate dose-response studies;
- Protocol should be pre-registered;
- High rate of missing data for aging animals.

Good practices that could be implemented in cancer research

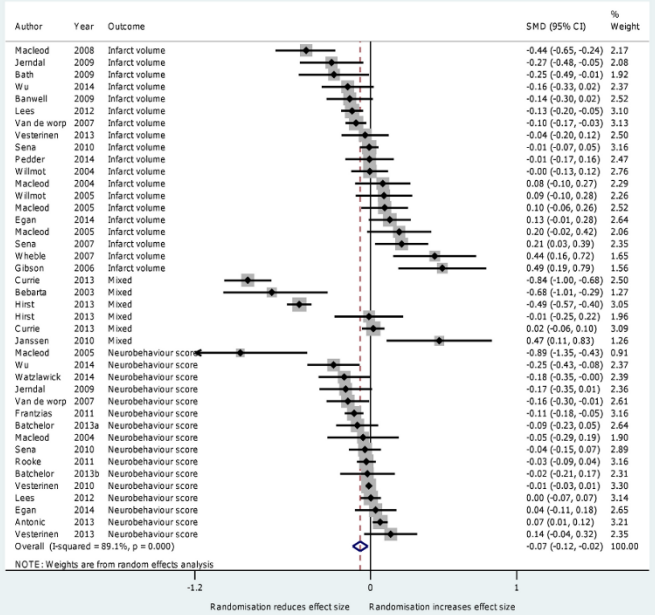
Randomization and Blinding

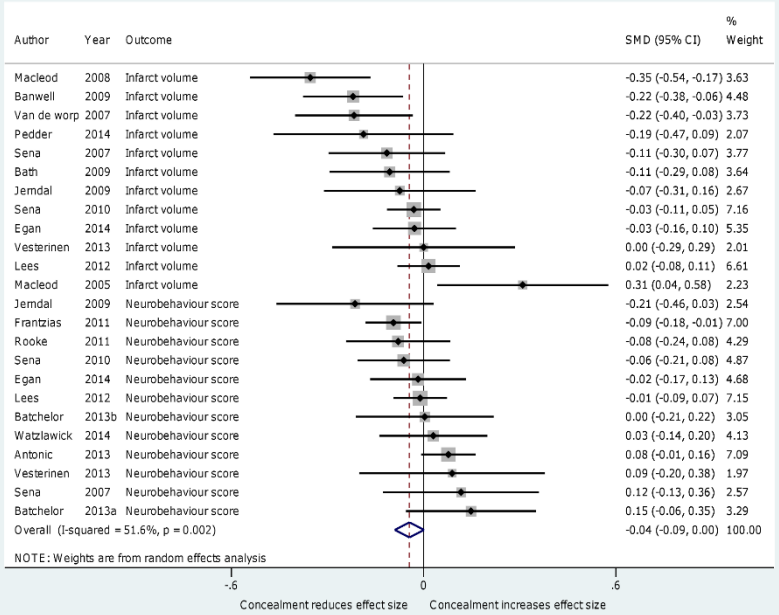
OPEN ACCESS Freely available online

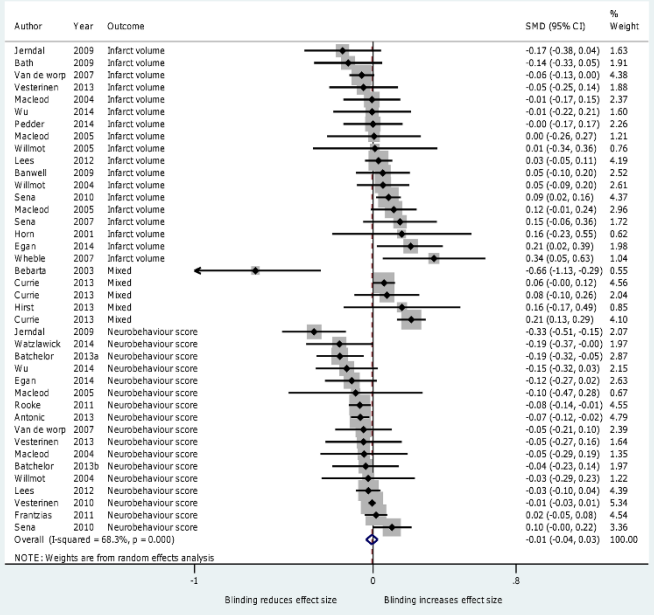
PLOS ONE

The Need for Randomization in Animal Trials: An Overview of Systematic Reviews

Jennifer A. Hirst[1*], Jeremy Howick[1*], Jeffrey K. Aronson[1], Nia Roberts[2], Rafael Perera[1], Constantinos Koshiaris, Carl Heneghan[1]







Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, Heneghan C. The need for randomization in animal trials: an overview of systematic reviews. PloS one. 2014 Jun 6;9(6):e98856

Stratification by sex

- In 2016, the National Institutes of Health (NIH) implemented a policy requiring investigators to consider sex as a biological variable;

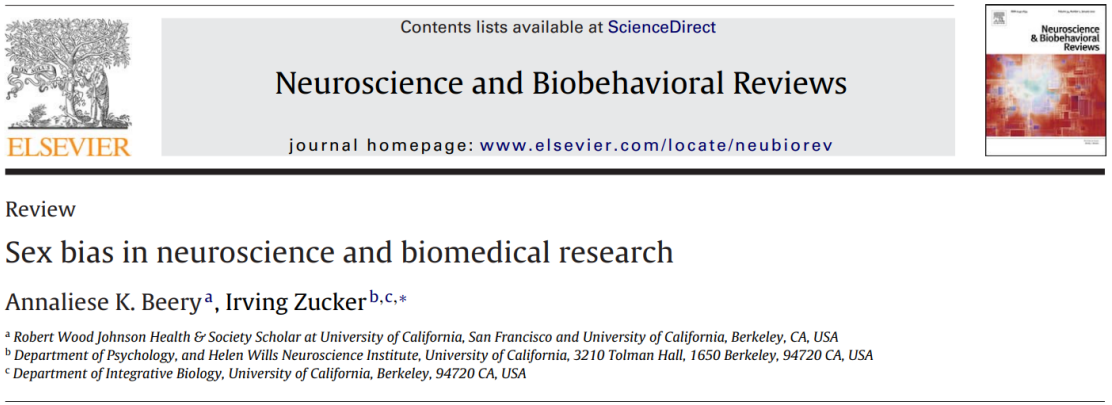
- NIH policy was a consequence of a series of reports calling for the inclusion of females in research and describing the limitations of sex-biased studies starting in the 1990s until early 2000s.

- The policy aimed to ensure equal representation of males and females in vertebrate research studies;

- It does not require investigators to power studies in order to determine sex differences nor does it ask investigators to analyze data by sex.

https://orwh.od.nih.gov/sex-gender/nih-policy-sex-biological-variable

Stratification by sex



Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Sex bias in neuroscience and biomedical research

Annaliese K. Beery[a], Irving Zucker[b,c,*]

[a] Robert Wood Johnson Health & Society Scholar at University of California, San Francisco and University of California, Berkeley, CA, USA
[b] Department of Psychology, and Helen Wills Neuroscience Institute, University of California, 3210 Tolman Hall, 1650 Berkeley, 94720 CA, USA
[c] Department of Integrative Biology, University of California, Berkeley, 94720 CA, USA

- In 2011, Beery and Zucker conducted a systematic review to quantify the extension of sex-bias across several research areas. Out of 841, only 28% (n = 232) articles had inclusion of both sexes such that 50% (n = 131) of them presented analysis by sex.

eLife FEATURE ARTICLE

META-RESEARCH

A 10-year follow-up study of sex inclusion in the biological sciences

NICOLE C WOITOWICH*, ANNALIESE BEERY AND TERESA WOODRUFF

- Woitowich et al. (2020) did a follow-up study including 720 articles from 9 research areas (including PloS Biology, Science, Nature among others).
- There was a large increase of sex-inclusive studies from 28% to 63%.
- However, there is no improvement on analyses by sex from 50% to 42%.

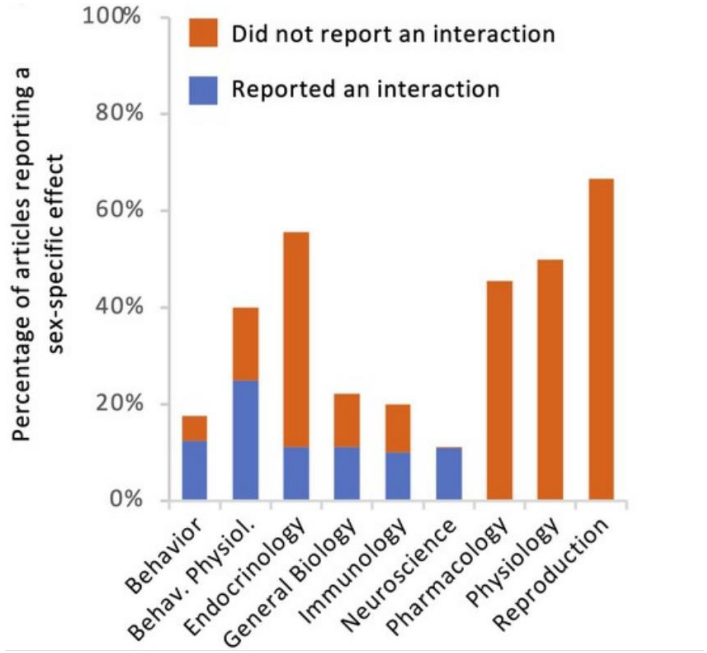
Beery, A.K. and Zucker, I., 2011. Sex bias in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, *35*(3), pp.565-572.
Woitowich NC, Beery A, Woodruff T. Meta-research: a 10-year follow-up study of sex inclusion in the biological sciences. Elife. 2020 Jun 9;9:e56344.

Stratification by sex

eLife RESEARCH ARTICLE

Reporting and misreporting of sex differences in the biological sciences

Yesenia Garcia-Sifuentes[1], Donna L Maney[1,2]*



- Garcia-Sifuentes and Maney (2021) evaluated 147 articles that had analyzed sex as a confounding variable, stratification and interaction among selected papers from Woitowich et al. (2020);
- Among those 147 articles, 92 (62%) planned an equal number of females and males in their studies;
- Among those 92 articles, 61 (67%) claimed sex differences but 40 (65%) did not test the interaction effect.
- Among those 40 articles,
  - 24 (60%) based their conclusions on the stratified analysis;
  - 12 (30%) based their conclusions on the comparisons between sex within a treatment arm;

Garcia-Sifuentes Y, Maney DL. Reporting and misreporting of sex differences in the biological sciences. Elife. 2021 Nov 2;10:e70817.

Introduction of controlled variability

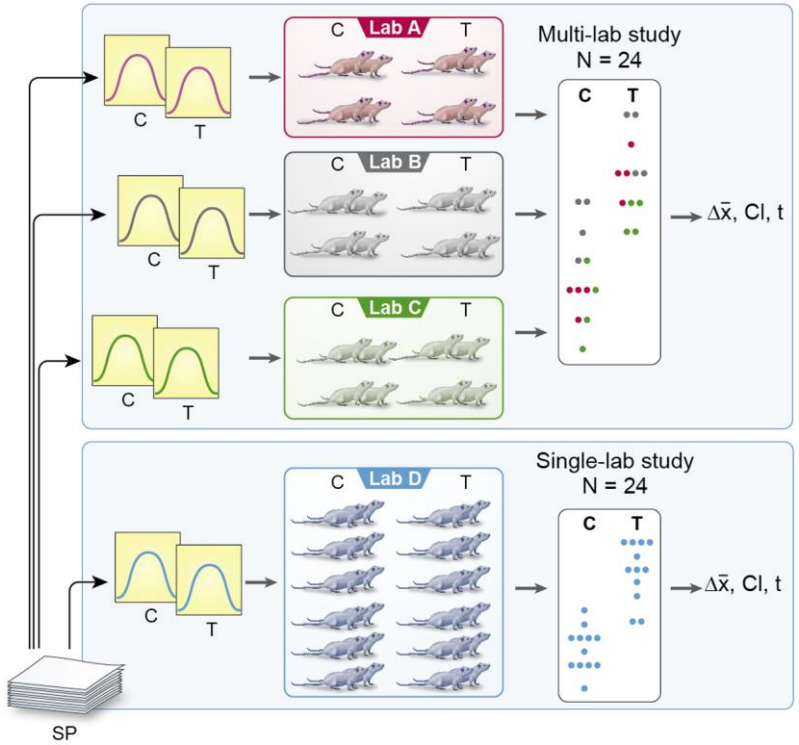
PLOS BIOLOGY

META-RESEARCH ARTICLE

Reproducibility of preclinical animal research improves with heterogeneity of study samples

Bernhard Voelkl[1], Lucile Vogt[1], Emily S. Sena[2], Hanno Würbel[1]*

- Although genetic and environment standardizations are considered gold standard yielding more homogeneous populations, such good practices might generate results that are too specific to standardized study conditions which leads to poor replicability/reproducibility;
- A simulation study was performed using data of 50 independent studies for stroke on the effect of therapeutic hypothermia on infarct volume in rodent models of stroke available on the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES);
- Multi-laboratory studies and potentially other ways of creating more heterogeneous study samples provide an effective means of improving the reproducibility of study results.



Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. PLoS biology. 2018 Feb 22;16(2):e2003693

Introduction of controlled variability

Systematic heterogenization for better reproducibility in animal experimentation

S Helene Richter

Lab Animal **46**, 343–349 (2017) | Cite this article

- Variability can be introduced in single laboratory studies with more than one mouse model/strain and mini-batches of experiments.

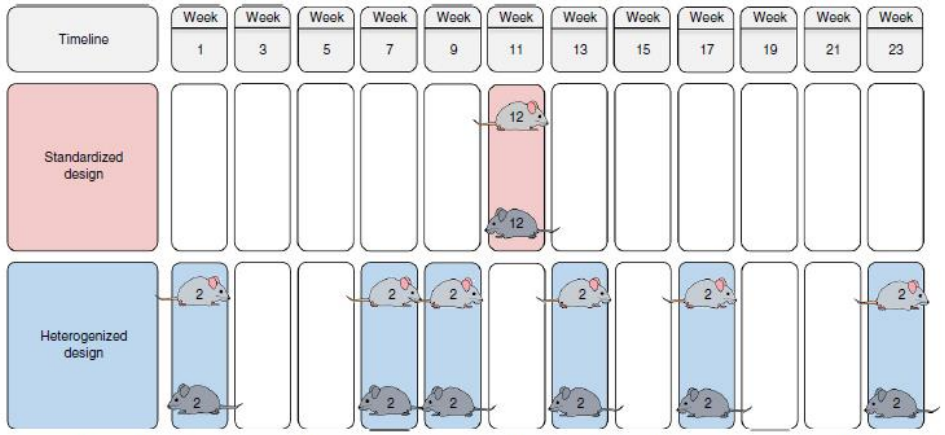


FIGURE 2 | Systematic heterogenization over time ("batch heterogenization"). Batch heterogenization aims to split experiments into small batches of animals that are tested some time apart (heterogenized design) instead of testing them at once in just one large batch (standardized design). Combining these "mini-experiments" in one big experiment is then assumed to increase representativeness of the whole study population, resulting in findings that are more reproducible between experiments and laboratories.

Richter SH. Systematic heterogenization for better reproducibility in animal experimentation. Lab animal. 2017 Sep;46(9):343-9

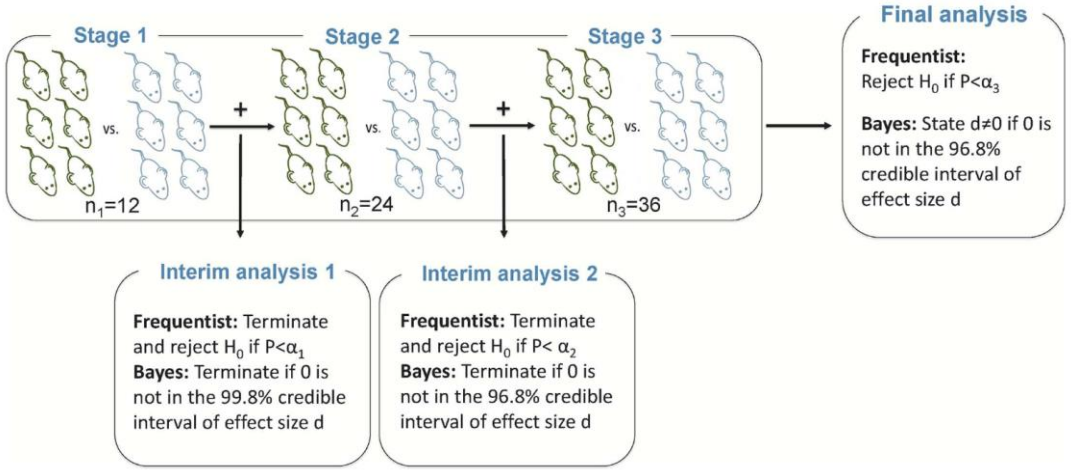
Adaptive Sample Size

PLOS BIOLOGY

PERSPECTIVE

Increasing efficiency of preclinical research by group sequential designs

Konrad Neumann[1☯], Ulrike Grittner[1,2☯*], Sophie K. Piper[1,2,3], Andre Rex[2,4], Oscar Florez-Vargas[5], George Karystianis[6], Alice Schneider[1,2], Ian Wellwood[2,7], Bob Siegerink[2,8], John P. A. Ioannidis[9], Jonathan Kimmelman[10], Ulrich Dirnagl[2,3,4,8,11,12]



- Group sizes in preclinical research are seldom informed by statistical power considerations but rather are chosen on practicability;
- Group sequential designs can offer higher efficiency than traditional methods and are increasingly used in clinical trials including futility or efficacy stopping rules.
- Sequential designs can lead to a substantial reduction in number of animals for some experiments allowing increased sample sizes to more promising experiments.

Neumann K, Grittner U, Piper SK, Rex A, Florez-Vargas O, Karystianis G, Schneider A, Wellwood I, Siegerink B, Ioannidis JP, Kimmelman J. Increasing efficiency of preclinical research by group sequential designs. PLoS biology. 2017 Mar 10;15(3):

Adaptive Sample Size

- The approach of sequentially collecting data, one measurement at a time, and stop when we have sufficient measurements, e.g. when the p-value drops below 0.05 seems very appealing when minimizing sample size is desired;

Sample sizes

For optogenetic activation experiments, cell-type-specific ablation experiments, and in vivo recordings (optrode recordings and calcium imaging), we continuously increased the number of animals until statistical significance was reached to support our conclusions. For rabies-mediated and anterograde tracing experiments, the selection of the sample size was based on numbers reported in previous studies. For optrode recordings, we first recorded a preliminary data set of six units from two mice. Based on analysis of this data set and given the success rate in finding identified GABAergic units, we predicted that about 20 units are sufficient to statistically support our conclusions.

Weber F, Hoang Do JP, Chung S, Beier KT, Bikov M, Saffari Doost M, Dan Y. Regulation of REM and non-REM sleep by periaqueductal GABAergic neurons. Nature communications. 2018 Jan 24;9(1):1-3

Adaptive Sample Size

Comment | Open Access | Published: 23 April 2019

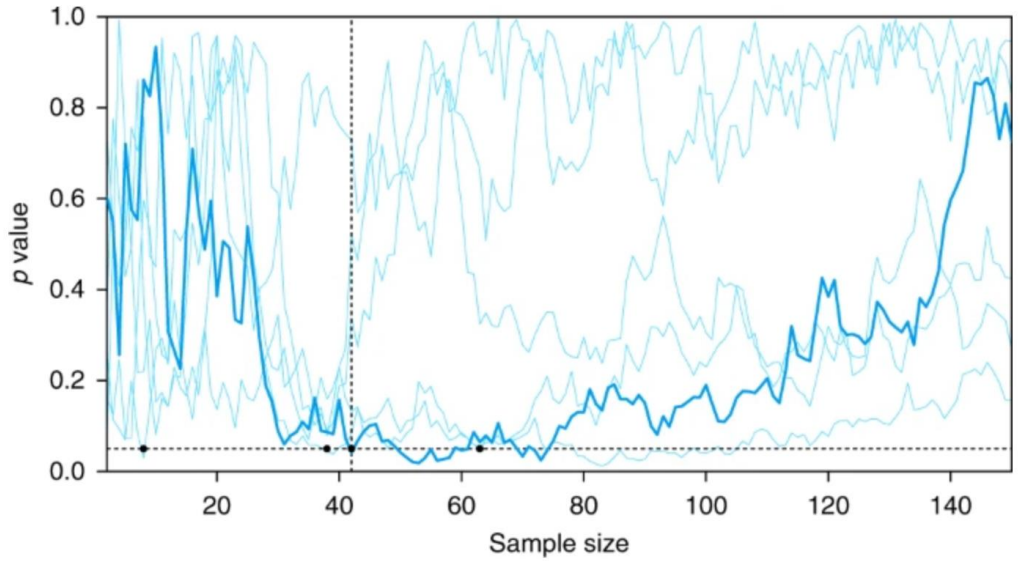
The problem with unadjusted multiple and sequential statistical testing

Casper Albers

Nature Communications **10**, Article number: 1921 (2019) | Cite this article

- Without adequate statistical methods, sequential testing increases the false positive rate;
- In various anonymous large-scale surveys, large numbers of researchers, active in various fields of research, have admitted to following this strategy at least once. Some of the findings include 36.9% of ecologists and 50.7% of evolutionary biologists.

Fig. 1



A computer simulation of sequential *p*-values when there is no effect. The thick line is the instance discussed in the text; the five thin lines represent independent simulations. The black dots indicate the first instance where one of the runs falls below the 0.05 level. Two of the runs don't reach 0.05 before $n = 150$

Albers C. The problem with unadjusted multiple and sequential statistical testing. Nature Communications. 2019 Apr 23;10(1):1921.

Opportunities of Interaction

- Design of experiments
- Strategies for randomization and blinding;
- Strategies to introduce controlled variability in experiments;
- Designs with adaptive sample size;
- Power considerations.

- Data analyses for in-vitro and in-vivo studies
- Statistical rigor to conduct test of pre-established hypotheses;
- Strategies to deal with missing data due animal death;
- Code for analysis available to share with publications;
- Reporting according to ARRIVE guidelines.

Biostatistics and Clinical Informatics Core at Tisch Cancer Institute

Biostatistics and Clinical Informatics Core

Aims of the Biostatistics team in Tisch Cancer Institute are to:

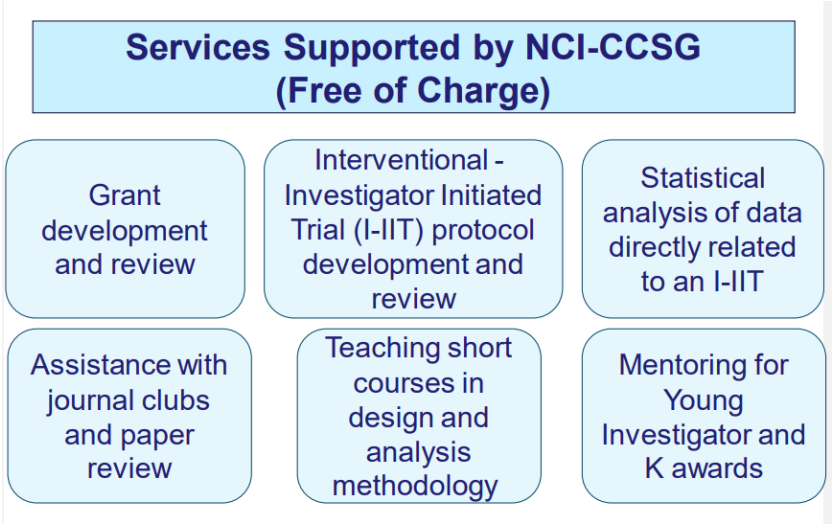
- Establish a scientific and administrative structure that supports investigators from a broad background and creates a collegial environment;

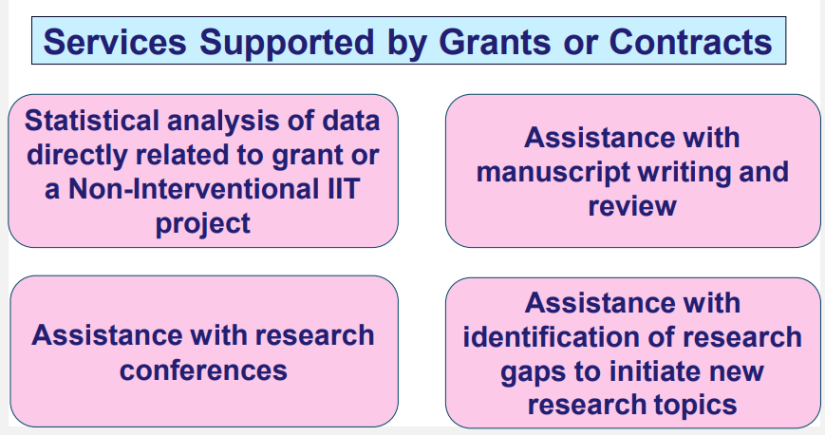
- Provide high-quality consultation in research design and biostatistical analysis;

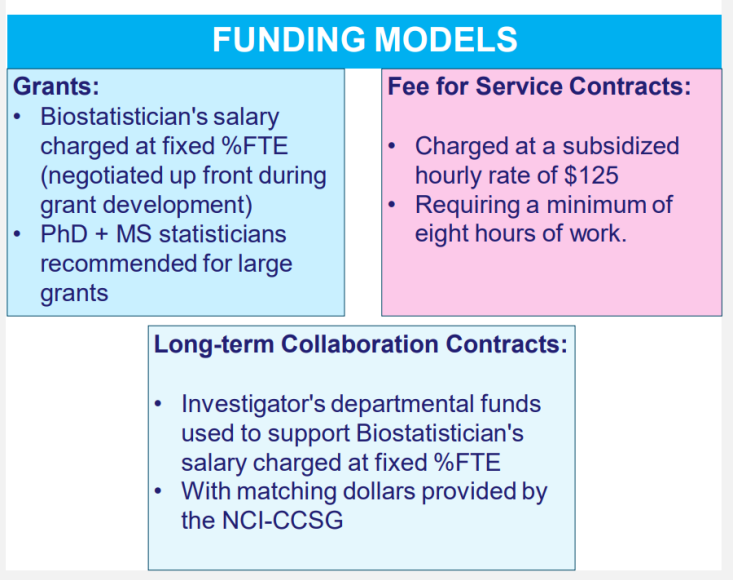
- Train laboratory and clinical investigators in the quantitative aspects of research;

- Support development of innovative statistical methods and promote application of novel analytical methods to collaborative projects.

Biostatistics and Clinical Informatics Core







Biostatistics Team

Madhu Mazumdar, PhD
Co-Director

Marcio Diniz, PhD
Co-Director

Erin Moshier, MS
Managing Director

John Mandeli, PhD
Associate Professor

Xiaoyu Song, PhD
Associate Professor

Deukwoo Kwon, PhD
Associate Professor

Himanshu Joshi, PhD
Assistant Professor

Seungjun Ahn, PhD
Assistant Professor

Francesca Petralia, PhD
Assistant Professor

Lewis Tomalin, PhD
Assistant Professor

Asem Berkalieva, MS
Senior Biostatistician

Weijia Fu, MS
Biostatistician II

Grace Van Hyfte, MS
Biostatistician II

Mayuri Jain, MS
Biostatistician I

Questions?

marcio.diniz@mountsinai.org

More information, visit:

https://labs.icahn.mssm.edu/tcibci/

Icahn School of Medicine at Mount Sinai